engineering and technology **tech** for the planet.

Evaluation of the CALPUFF Air Dispersion Model As Applied to Assess Class I SO_2 Increment Status in North Dakota

Prepared for:
The Environmental Law Group
10 Second Street Northeast, Suite 114
Minneapolis, Minnesota

Prepared by:
Earth Tech, Inc.
196 Baker Avenue
Concord Massachusetts

May 20, 2002

TABLE OF CONTENTS

LIST OF FIGURES

1.0 INTRODUCTION

Earth Tech has reviewed the application of the CALPUFF model for the assessment of SO_2 increment in Class I areas in western North Dakota and eastern Montana. This review is concerned primarily with technical issues relating to the application of CALPUFF and its companion meteorological model, CALMET, including the evaluation study performed by the North Dakota Department of Health (NDDH), which was used to select and justify alternative options and parameter settings for CALMET/CALPUFF for this application.

CALPUFF is a non-steady-state modeling system that includes meteorological and geophysical data processors, a meteorological model, a puff-based dispersion model, and postprocessing modules.

Earth Tech's Atmospheric Studies Group (ASG) provides research and consulting services in the environmental and physical sciences. The group specializes in air quality model development, atmospheric boundary layer research, air quality permitting and licensing, and regulatory consulting. The group's principals and supporting staff are highly qualified scientists and engineers with many years of professional experience. We have developed and evaluated numerous meteorological and dispersion models for both public agencies and private industries. These ongoing efforts serve to maintain our position at the leading edge of modeling and scientific technology and better enable us to assist clients in planning for and solving environmental problems. In addition, the group's commitment to developing and using the latest scientific advances is supported by its strong staff credentials in managing and conducting environmental research projects and by a corporate recognition that the command of "state-of-the-art" technology is essential in resolving today's complex environmental issues.

As the developers of the CALPUFF and CALMET models, Earth Tech is thoroughly familiar with the choices of options and inputs that confront the user. Earth Tech continues to refine and enhance the CALPUFF modeling system, in order to incorporate scientific advances, to accommodate new or improved databases for input information, and to provide greater flexibility for model users. On a regular basis, Earth Tech offers an intensive three-day training course that includes hands-on application of the models and provides an introduction to their technical basis.

2.0 MODEL SELECTION

The North Dakota Department of Health (NDDH) and EPA Region 8 selected the CALPUFF model to assess SO_2 increment consumption for Class I areas in western ND and eastern Montana. The proposed model application involves impact assessment for source-receptor distances ranging from a few kilometers (km) up to about 300 km. Federal guidance relating to air quality modeling[1] distinguishes between "near field" and "long range transport" modeling applications; a cutoff distance of 50 km is used to delineate the maximum distance at which near field techniques should be used. The present Class I increment application is clearly in the long-range transport category.

Under the Clean Air Act, Congress required EPA to "specify with reasonable particularity each air quality model or models to be used under specified sets of conditions for purposes of [PSD]." 42 U.S.C. 7475(e)(3). To meet this requirement, EPA has set forth approved models in its Guideline on Air Quality Models included in 40 CFR Part 51, Appendix W. At the 7th Modeling Conference[2], the EPA Office of Air Quality Planning and Standards (OAQPS) proposed the CALPUFF[3] model as the recommended model for long-range transport applications. The EPA recommendation of CALPUFF is based in large part on the recommendations of the Interagency Workgroup on Air Quality Models (IWAQM). IWAQM, comprised of representatives from federal and state agencies involved in regulating air quality, performed a comprehensive review and evaluation of modeling techniques suitable for long-range transport (LRT) applications. IWAQM considered model performance for a range of CALPUFF model options and input parameters, and the IWAQM Phase 2 report[4] includes recommendations relating to parameters and options for applying CALPUFF and its companion CALMET meteorological model. EPA-OAQPS, in recommending CALPUFF as an Appendix W model, incorporated most of the IWAQM recommendations concerning options and parameter settings for CALMET and CALPUFF.

The Federal Land Managers' Air Quality Related Values Workgroup (FLAG) has also developed recommendations for air quality modeling to assess impacts on Class I areas[5]. The recommendations specifically address modeling procedures for assessing impacts on visibility, vegetation (ambient ozone) and acidic deposition (sulfate, nitrate), but the guidance is closely related to recommendations for LRT applications, since impact assessment for Class I areas usually involves distances beyond 50 km. CALPUFF is the primary model recommended by FLAG for the assessment of air quality related values in Class I areas. FLAG recommendations also reference the IWAQM studies.

Since the 7th Modeling Conference, EPA has not published an updated version of the Guideline, so CALPUFF is not approved as a recommended ("Appendix W") model for long-range transport applications. At the present time, there is no long range transport model currently approved in EPA's Guideline on Air Quality Models.

---

[1] *Guideline on Air Quality Models*, 40 CFR Part 51, Appendix W

[2] Proposed Rules, Federal Register, April 21, 2000, page 21506

[3] Scire, J.S., D.G. Strimaitis and R.J. Yamartino. *A Users Guide for the CALPUFF Dispersion Model (Version 5.4), 2000*

[4] *Interim Recommendation for Modeling Long Range Transport and Impacts on Regional Visibility*, EPA Publication No. EPA-454/R-93-015, 1993

[5] Federal Land Managers' Air Quality Related Values Workgroup (FLAG) Phase I Report, December 2000

3.0 APPLICATION OF CALPUFF BY NDDH AND EPA REGION 8

CALMET and CALPUFF were applied with a 10 km grid scale, for an area that extends 640 km east-west by 460 km north-south. For CALMET, the meteorological inputs supplied to the model include hourly surface measurements from 25 stations located in or near the modeling grid, plus twice-daily upper air measurements from six stations (two inside of the modeling grid). Precipitation data from 96 observing stations located in or near the modeling grid was used.

The application of CALPUFF and CALMET by NDDH and EPA Region 8 for the Class I increment study did not consistently follow the recommendations of EPA-OAQPS (proposed) or IWAQM concerning model options and parameter settings. The technical decisions relating to options and parameter settings were made primarily by NDDH, based on a limited model performance/sensitivity study, (see item 3 below); EPA Region 8 then adopted the NDDH approach. EPA Region 8 also ran CALPUFF with the EPA-recommended ("IWAQM") settings. When using observations to select model parameters and options to achieve improved model performance, it is important to recognize the limitations of available measurements, and precautions should be taken to avoid "model tuning" (i.e., calibration via sensitivity testing) to achieve apparent agreement between predictions and observations.

A modeling protocol represents the most effective mechanism to ensure that the modeling approach is technically sound and consistent with regulatory guidance. It also provides a control mechanism to document that the modeling approach has been defined in advance, based on technical and regulatory criteria, and was not modified for convenience to achieve desired results. The Class I increment modeling for ND and eastern MT has had a somewhat circuitous history. The EPA Region 8 modeling approach is based on the NDDH 2001 modeling study[6], which evolved out of the earlier NDDH 1999 modeling study[7]. The NDDH 2001 approach is based on the NDDH model performance assessment. The protocol prepared by NDDH in 2000[8] for the 2001 NDDH application is fairly cursory and made no provision for a performance assessment study. No planning document for the model evaluation study has been released by either agency, and results have only been published for the final model configuration selected by NDDH. Aside from a few remarks about model performance using "IWAQM" settings, no description of the alternatives that were evaluated, the sensitivity of model predictions to different options and parameter settings, or model performance results for any tested alternatives have been published. Comments by EPA's Office of Air Quality Planning and Standards (OAQPS)[9] on the modeling by EPA Region 8 also suggested that other model options be considered. The response of EPA Region 8 or NDDH to the suggestions by EPA-OAQPS is not known.

---

[6] North Dakota Department of Health, *Evaluation of CALPUFF Model Performance Using Year 2000 Data,* November 2001

[7] North Dakota Department of Health, *CALPUFF Class I Area Analysis for Milton R. Young Generating Station,* May 1999

[8] North Dakota Department of Health, *Class I Increment Analysis for SO_2 Modeling Protocol*, 2000

[9] OAQPS Review of the Region VIII January 2002 Draft report: Dispersion Modeling Analysis of PSD Class I Increment Consumption in North Dakota and Eastern Montana; (attachment titled 'Review of R8s Report on ND Modeling' to email from Melissa McCullough, et al, January 25, 2002, USEPA Office of Air Quality Planning and Standards.

4.0 NDDH PERFORMANCE EVALUATION STUDY

Model performance results for CALPUFF were published as part of the 2002 NDDH modeling report[10]. Model predictions for calendar year 2000 were compared to observed SO_2 concentrations at two monitoring sites. The locations of monitors and major sources of SO_2 emissions are shown in Figure 1. The monitor located at the South Unit of Theodore Roosevelt National Park (TRNP) provides SO_2 measurements representative of that Class I area, while the Dunn monitor is located about 60 km east of TRNP (Distances from the Dunn monitor to the group of power plants east of that monitor site range from about 50 km to 105 km, while distances from the South Unit monitor to those power plants range from about 125 to 175 km).

Results of the limited comparison show predicted peak 3-hour average and 24-hour average concentrations for the year are within a factor of two of observed concentrations at both monitor locations. Results for the TRNP South Unit monitor show a consistent bias toward overprediction of peak 3-hour and 24-hour average concentrations, while the results for the Dunn monitor show little or no bias between predictions and observations. With comparisons based on only one year of data from two sites, it is not possible to establish a clear pattern of model performance applicable to the Class I areas of concern. What data exists in the Class I area suggests an overprediction bias at the South Unit, but additional performance evaluation data is needed.

The description of the performance evaluation in the 2002 NDDH report indicates that CALMET/CALPUFF was tested with a variety of options and parameter settings, but this testing has not been described in any published documents, and no data pertaining to the evaluation study has been released by either NDDH or EPA Region 8. It is therefore unclear how NDDH selected the "final" model options and settings, or whether the chosen settings provided better performance at the South Unit monitor than any of the other alternatives considered.

A diagnostic evaluation is key component of performance testing. Diagnostic analysis looks for characteristic patterns associated with peak observed concentrations, and then examines whether peak predictions follow similar patterns. For example, peak observed concentrations may show distinct seasonal or diurnal patterns, or may be associated with specific types of meteorological conditions. The goal of diagnostic analysis is to assess whether the model is correctly accounting for the processes that lead to high concentrations. EPA guidance on model performance testing recommends diagnostic analysis as a basic component of performance evaluation[11,12]. The NDDH report does not describe any such diagnostic analysis.

A comparison of the seasonal patterns of observed and predicted 24-hour average peak values illustrates the type of diagnostic analysis that is needed. Figure 2 compares the frequency of days with peak 24-hour increment predictions in the TRNP South Unit exceeding 5 $\mu g/m^3$, (EPA Region 8 modeling results for 1990-1994, CALPUFF with NDDH settings), versus peak 24-hour observed concentrations at the South Unit monitor exceeding 6 $\mu g/m^3$, for 1998-2001. As this figure illustrates, the majority of peak observed high impacts occur in the winter, while only 4 of 34 peak predictions occur in the winter. The high

[10] North Dakota Department of Health, *CALPUFF Analysis of Current PSD Class I Increment Consumption in North Dakota and Eastern Montana Using Actual Annual Average SO2 Emission Rates,* Draft, April 2002
[11] D.G. Fox, *Judging air quality model performance. A summary of the AMS workshop on dispersion model performance, Woods Hole, MA, September 1980,* Bulletin American Meteorological Society **62**:599-609, May 1981
[12] EPA-OAQPS, *Interim Procedures for Evaluating Air Quality Models (Revised)*, EPA-450/4-84-023, 1984

frequency of peak observed days during winter months is characteristic of observations at the South Unit from 1980 through 2001.

Having evaluated performance for calendar year 2000, NDDH and EPA could (and should) also have performed increment analysis using the 2000 data set (plus other years, if necessary). Since model performance was only tested for a single year, it is unclear whether performance results are representative of peak events in other years (e.g., 1990-1994). Increment analysis using 2000 data would at least indicate whether peak predictions for 2000 are similar to those for 1990-1994.

It should be noted that the NDDH performance evaluation compared peak observed and predicted peak concentrations for 3-hour and 24-hour averages at each monitoring site, *unpaired in time*. This is the standard procedure for testing model performance, in recognition that dispersion models are generally unable to predict the exact time and location where peak impacts will occur. (A very small error in wind direction, for example, can shift the location of predicted impacts.) This "lack of skill" for predicting impacts paired in time has important implications for the use of models to assess PSD increment consumption. In general, any dispersion model is better suited for predicting the magnitude of peak concentrations that will occur at a given location over a large number of events (e.g., one or more years) than for predicting the concentration for a specific event.

Dispersion models are therefore better suited to estimating PSD increment consumption by computing the peak (e.g. highest second-highest, or H2H) *current* concentration and the peak *baseline* concentration, and then taking the difference between these peak predicted values, independent of time, as opposed to estimating the *difference* between current and baseline concentrations event-by-event. The NDDH method of computing increment consumption corresponds to the first approach, i.e., peak current and baseline concentration values are compared on a time-independent basis. The standard EPA procedure for assessing increment consumption, by contrast, uses the model (in this case, CALPUFF) to predict the difference between current and baseline impacts at each receptor, event-by-event.

5.0 EPA PERFORMANCE EVALUATION

The EPA Region 8 modeling report makes no mention of any performance evaluation that they performed. EPA explicitly relied on the NDDH performance evaluation to assess the performance of CALPUFF with the NDDH settings. No corresponding evaluation of CALPUFF with IWAQM settings was performed. The NDDH performance evaluation report, however, stated that "*Changing all control file settings to IWAQM-recommended values, for example, would likely move some predicted-to-observed ratios outside of the factor-of-two window.*"[13]

[13] North Dakota Department of Health, *Evaluation of CALPUFF Model Performance Using Year 2000 Data*, November 2001

6.0 OPTIONS FOR ADDITIONAL SENSITIVITY ANALYSIS

In light of the prediction bias demonstrated in the limited NDDH evaluation for the South Unit monitor, and with additional SO_2 monitoring data now available from the North Unit as well as the South Unit, a more comprehensive, systematic analysis of model sensitivity and model performance to evaluate a broader range of options for the application of CALMET and CALPUFF is clearly necessary. At a minimum, options that need to be considered include the following:

- Apply CALMET in conjunction with a prognostic mesoscale meteorological model, such as the Penn State MM5 model. EPA, in proposing CALPUFF as a Guideline model, encouraged the use of output from a prognostic mesoscale model for long range transport applications[14]. MM5 output may be particularly valuable in a comparatively data-sparse region such as western North Dakota.

- Apply CALMET and CALPUFF with a smaller grid scale, such as 2-5 km, and with additional layers to improve vertical resolution. A finer grid would offer even greater advantages if CALMET were run using MM5 meteorological fields. While a finer grid will impose a greater computational burden, grid resolution may affect the performance of model features such as puff-splitting.

- Apply and document CALMET and CALPUFF performance with a variety of other technical options, including the EPA-recommended "IWAQM" settings.

- Expand the evaluation database to include 2001 SO_2 measurements from the South and North Unit monitors.

- Conduct the evaluation based on meteorological data for the same years for which modeling will be conducted.

- Perform a diagnostic analysis of observed and predicted peak SO_2 concentrations, for modeling alternatives to compare the characteristics of peak observed and predicted events (season, time of day, meteorology).

[14] Proposed Rules, Federal Register, April 21, 2000, page 21527

7.0 PREDICTION BIAS AT SOUTH UNIT

The NDDH performance evaluation shows peak 3-hour average predictions for the South Unit that exceed observations by a factor of 1.5 to 1.8 (for 2000). In the increment analysis performed by EPA Region 8, the highest second-highest (H2H) increment prediction for the South Unit is 31.4 μg/m^3, which exceeds the Class I increment by only a factor of 1.27. When prediction bias is considered, these results indicate that the 3-hour average Class I increment is not exceeded at the South Unit.

For 24-hour averages, the NDDH performance evaluation shows that peak predictions exceed observations by a factor of 1.3 to 1.7. For four of five years (1991-1994), the EPA Region 8 modeling analysis gives H2H 24-hour average increment predictions of 8.4 μg/m^3 or less, or 1.6 times the Class I increment, comparable to the model bias demonstrated for 2000. The H2H prediction for 1990 for the South Unit is 12.8 μg/m^3, roughly 2.6 times the Class I increment, and much higher than the bias found for 2000.

The increment predictions for the South Unit, which are as large as the exceedances predicted for any of the Class I areas, based on the EPA modeling, can be attributed in large part to the prediction bias demonstrated by the NDDH evaluation. These results illustrate how prediction bias, even within the factor-of-two range, can result in the diagnosis of increment exceedances where none exist.

Although no performance evaluation was reported for CALPUFF with IWAQM settings, it is clear that overprediction bias is more extreme for this option. The *increment* predictions for the South Unit reported by EPA using CALPUFF with IWAQM settings (and with NDDH settings) exceed the highest 3-hour and 24-hour average SO_2 concentrations observed at the South Unit monitor over the past ten years, even though the observed concentrations include background and baseline contributions.

TABLE 1

COMPARISON OF MEASURED CONCENTRATIONS AND PREDICTED INCREMENT CONSUMPTION

Measured Concentrations (including baseline, background) 1992-2001		EPA CALPUFF Predicted Increment Consumption 1990-1994 Regulatory Default		EPA CALPUFF Predicted Increment Consumption 1990-1994 NDDH Options	
(a) 24-Hour Average SO_2, TRNP South Unit (μg/m^3)					
Max	H2H	Max	H2H	Max	H2H
10.4	9.9	22.4	18.6	15.3	12.8
(b) 3-Hour Average SO_2, TRNP South Unit (μg/m^3)					
Max	H2H	Max	H2H	Max	H2H
39.3	30.6	61.5	45.1	36.4	31.4

8.0 CONCLUSIONS

Documentation of the sensitivity analysis conducted by NDDH to select an alternative modeling approach that departs from EPA recommendations is incomplete. The technical basis for choosing specific model options and parameter values is not adequately explained, the other options that were evaluated are not described, and model performance for the recommended "regulatory option" is not documented. EPA has not conducted, or at least not reported, any performance evaluation or validation study of model performance specific to its application of the model in North Dakota.

Model performance results at the one monitor representative of Class I area impacts show systematic overprediction bias for peak concentrations. Model results also are inconsistent with actual monitored SO_2 concentrations. A more complete and comprehensive model evaluation/sensitivity study that: (1) uses additional SO_2 measurements from both the South Unit and North Unit collected since 2000; (2) evaluates model performance and assesses increment consumption for the same year(s); and (3) assesses whether peak predictions and observations occur for similar events, is needed to reach conclusions regarding model validity for purposes of evaluating North Dakota increment consumption.